图书情报工作
第64卷 第5期 2020年3月

基于引文曲线拟合的新兴技术主题的突破性预测*
——以干细胞领域为例

■ 曹艺文[1,2] 许海云[1,3] 武华维[1,2] 罗瑞[1,2]
[1] 中国科学院成都文献情报中心 成都 610041 [2] 中国科学院大学经济与管理学院图书情报与档案管理系 北京 100190
[3] 中国科学技术信息研究所 北京 100038

摘 要：[目的/意义]通过对新兴技术主题的引文曲线进行拟合分析，提炼和归纳引文曲线的主要类型和特征，以期能为微观层面突破性创新主题预测方法的研究提供参考。[方法/过程]首先提出利用引文曲线追踪感知新兴技术主题形成突破性创新的假设，并提出跃迁指数的概念和计量方法，从知识跃迁和持续增长两个维度总结识别突破性创新的标准并构建突破性识别的方法模型。其次利用新兴技术主题发文的时序被引数据构建引文曲线并进行拟合分析，实现对不同主题包含的引文曲线的类型分类，提炼引文曲线的主要特征进行突破性预测。[结果/结论]新兴技术主题引文曲线的特征包括：近期持续增长；近期持续下降；生命周期短；相近年份变化趋势一致。根据突破性创新的识别标准，如果引文曲线跃迁次数多、跃迁幅度大，最高被引峰值出现时间较晚，并且在近期保持高被引且持续稳定或快速增长，则新兴技术主题具有成为突破性创新的潜力。结合领域专家对预测结果的评估以及不同新兴技术主题的研究进展，验证了利用引文曲线可以有效预测新兴技术主题的突破性。

关键词：突破性创新 技术预测 引文曲线 新兴技术主题
分类号：G250
DOI：10.13266/j.issn.0252-3116.2020.05.011

1 引言

随着知识全球化的发展与国际竞争的加剧，世界各国的科技创新已经进入空前活跃的时期，如何从不断涌现的新兴技术主题中识别和预测具有颠覆性和突破潜力的技术，对于国家和企业的技术战略布局至关重要。突破性创新的创新强度较高，往往是形成产业核心技术的基础[1]。尽早识别突破性创新对于国家或企业缩短创新周期，解决行业技术壁垒，构建新兴技术产业集群，促使突破性技术大量涌现从而提高我国的科技竞争力意义重大。但是由于突破性创新采用的是新技术，脱离原有的技术轨道，因此突破性创新相较于渐进性创新而言未来发展的不确定性更大，规律难以把握，识别难度更大。新兴技术（emerging technology）指的是新颖度高、相对增长速度快，具有一定的连贯性，可能会对社会经济产生重大影响的技术，但在最初兴起阶段具有较高的不确定性和模糊性[2]。新兴技术可能会对市场格局和行业发展产生持续性影响，甚至演化成能够改变已有技术范式和技术轨迹的突破性创新，也可能仅仅昙花一现。新兴技术主题虽然具有发展成为突破性创新的潜力，但并不代表它在未来一定有研究价值和前景[3]。因此，从新兴技术主题中识别潜在的突破性创新主题意义显著。

目前已经形成了一些突破性创新的识别方法，但现有研究多是对已经发生的突破性创新进行定性或定量分析。其中，定性分析方法仍然是发现技术发展趋势的重要手段，但多学科交叉融汇使得专家智慧和知识结构存在一定局限[4]，方法的主观性强且实施过程复杂，开展过程中要消耗大量的人力、物力和财力，耗时长，并且实施困难较大。定量预测方法以技术演化分析和专利分析为主，技术演化法需要事先构建识别指标，由于不同技术领域内突破性创新的发展模式存

* 本文系国家自然科学基金项目“基于科学-技术主题关联分析的创新演化路径识别方法研究”（项目编号：71704170）和中国科学院十三五信息化项目“面向干细胞领域知识发现的科研信息化应用”（项目编号：XXH13506）研究成果之一。
作者简介：曹艺文（ORCID：0000-0002-2439-1101），硕士研究生；许海云（ORCID：0000-0002-7453-3331），副研究员，博士，通讯作者，E-mail：xuhy@clas.ac.cn；武华维（ORCID：0000-0001-6969-407X），博士研究生；罗瑞（ORCID：0000-0002-5234-8901），硕士研究生。
收稿日期：2019-05-15 修回日期：2019-09-20 本文起止页码：100-113 本文责任编辑：杜杏叶

在差异，难以提炼出普适性的评估指标。专利分析法则会受发明人引用动机的影响存在局限性，即发明人为保证技术的新颖性可能回避相似专利。

本研究对尚无定论的新兴技术进行突破性预测，属于对新兴技术主题的前瞻性分析。研究拟从引文曲线可表征知识跃迁和持续增长两方面，构建新兴技术主题突破性识别的方法模型，并以干细胞领域为例，对时间切片形式的引文曲线拟合分析，从引文曲线自身的特征出发，识别和预测具有成为领域内重大关键技术的突破性技术，以期提高和完善突破性创新定量识别的理论和方法体系。

2 研究现状

2.1 突破性创新内涵

突破性创新(radical innovation)最早来源于熊彼特谈到的"创造性的破坏"[5]。真正意义上的突破性创新研究兴起，是在G. Dosi的经典论文《技术范式与技术轨道》发表之后，他将突破性技术创新和渐进性技术创新统一到一个理论框架之内[6]。现有对于突破性创新的大多数研究集中在：新产品、新工艺、新技术、全新的科学知识基础、提升产品性能、节约成本以及改变市场竞争结构等方面。突破性创新一方面关注技术本身的突破性特征，另一方面从创新产生的实质影响力进行定义。

从技术突破出发，突破性创新是一种全新的技术，建立在一整套全新的工程和科学知识的基础之上[7]，会颠覆现有的技术体系和认识[8]，需要全新的科技知识和资源，是一种高度非连续性的、具有革命性本质的创新，也是一种能力破坏型的创新[9]。从实质影响力出发，突破性创新能够提升产品性能、创造新产品、改变市场结构或竞争态势。一方面，突破性创新可以使现有的产品、工艺或服务具有前所未有的性能或虽具有相似的特征但性能实现大幅提高且成本降低[10]。突破性创新还能够创造新市场、新产品、新工艺、新产业、新的业务模式[7,11]。另一方面，突破性创新会淘汰现有技术和产品[9]，改变产业竞争结构，提高绩效，节约成本[11]，改变市场消费模式[12]，更好地满足顾客的需求[13]，对市场规则、竞争态势、产业版图具有决定性影响，甚至引起产业重新洗牌[14]。

本文的突破性创新侧重技术创新的动态性、不连续性和新颖性特征，关注的是创新本身所依托的全新的科学知识基础和产生重大科学突破的潜力，而非对现有技术的渐进性改进。

2.2 已有的突破性研究的定量识别方法

国内外学者从不同的角度展开了突破性创新识别和预测的一系列研究，主要包括专利分析视角、技术演化视角、主题突变视角、网络结构视角以及睡美人文献视角。

2.2.1 专利分析视角

利用专利识别突破性创新主要包括专利被引分析和专利语义差异两种方法。专利被引方面，W. Schoenmakers和G. Duysters使用半导体领域的数据验证高被引专利更可能催生突破性创新[15]。K. B. Dahlin和D. M. Behrens利用专利引用结构的差异识别破坏性创新，并通过网球拍相关的专利验证了引用结构差异越大，越有可能产生突破性创新[16]。张金柱从基础研究对技术创新的影响角度出发，以专利直接和间接引用的科学论文为基础，通过专利引用的科学知识突变对可能产生突破性创新的时间、研究主题和学科分类组合进行识别[17]。H. Small认为识别科学发现的方法受多重因素的影响，共被引分析就是识别多重发现的有效方法，多重(复合)是一个重要现象，其实质反映的是科学发展中，科学共同体之间的竞争与合作[18]。专利语义差异方面，研究人员主要通过专利分类聚类及其突变、专利主题聚类及其突变、专利权人合作以及跨界合作变化等，从突破性创新的源头对其进行识别、预警和预测。例如，J. Yoon和K. Kim基于SAO语义向量计算专利文本相似度寻找离群专利，这些离群专利代表了新型的突破性技术创新[19]。

2.2.2 技术演化视角

技术演化视角从技术轨道的转换论观点出发，寻求在一个新的技术轨道还没有成型时，判断它的转换期。A. Sood和G. J. Tellis发现，部分行业技术轨道在经历了长期稳定发展后出现了随机性的跳跃[20]，P. Anderson和M. L. Tushman发现行业单一技术性能的演进同样存在随机性的跳跃[21]。突破性创新通常出现在新旧技术轨道的更迭期并改变原有技术轨道，而渐进性创新则是在原有技术轨道上的性能改进[20]。但技术轨道需要事先定义哪些性能指标表征突破性，某一性能的指标不足以完整地描述技术演化，其剧变也不等同于突破性创新的发生[17]。G. J. Tellis进一步指出技术性能的改进并不单纯遵循S曲线，技术演化过程中突破性技术的涌现充满了随机性和偶然性，因此技术轨道S曲线并不是很好的预测工具，不应该作为制定战略的基础[22]。

2.2.3 主题突变分析

库恩认为科学进步的轨迹是跳跃式的[23]。主题

突变分析关注的是主题在某一时刻的重大变化程度，而不受主题关注阈值、领域研究热度等外界因素的影响。随着时间的推移，突变主题可能演变为研究热点，也可能消失或转换为普通主题[24]。目前对主题突变的研究主要是基于突发词和离群数据的分析。J. Kleinberg 提出了突发监测算法，根据词频变化率找出各时间段中的突发词，通过分析突发词状态的变化来发现新的研究热点[25]。张金柱基于引用知识的差异程度衡量引用科学知识的突变程度，进行了科技领域的突破性创新识别研究[26]。王莉亚认为模式异常、具有明显离群特点的数据往往能导致新知识的发现，并将离群数据概念引入到主题演化分析中[27]。

2.2.4 网络结构或转折点识别突破

新技术出现在由其他技术共同组成的技术环境中[28-29]，一项新技术带来的效应，不仅要看其自身影响，还要分析新技术如何适应现有的轨迹。R. J. Funk 通过网络分析方法挖掘新技术对以往技术以及整个技术轨道的影响[30]。在技术变革的领域，利用定量方法区分这些创新是脱离还是增强已建立的技术轨迹，将使实证研究更好地匹配基础理论，并可能促进新概念的发展。

在网络动态发展视角下，新信息的进入将影响结构的稳定性[31]，因此，通过对网络结构变化程度的测度可以识别知识创新对原有知识的影响力，其中引起网络结构巨大变化的信息可能是潜在的突破性创新。C. Chen 等通过中心度变化率 Δcentality 和模块化变化率 Δmodularity 探测具有潜在重要性的新出版物，分析发现 Δmodularity 值高的节点论文的得分取决于是否在网络的聚类间增添连接，即这篇论文是在原本没有关联的知识模块之间创建了一个桥或边界融合连接，并认为潜在的重要科学发现多出现在不同领域科学知识的交叉融合处[32]。

2.2.5 睡美人文献识别突破

超前性研究往往超前于现有认知领域，变革性研究由于颠覆现有研究范式，科学共同体不知其存在或意识不到其潜在的知识价值，对此保持较大的心理距离，从而低估其知识价值[33-34]。因此，与渐进性创新相比，突破性创新更容易受到忽视。科学史表明，一些重大的科学发现和创新成果没有被当时科学共同体的其他成员所及时接受而受到忽视或抵制，多年后才被人们重新认识，这类成果的文献称为科学中的“睡美人”(sleeping beauties)[35]。科学和技术领域都有延迟承认现象，在科学中表现为睡美人文献，在技术中表现为睡美人专利(sleeping patents)[36]。杜建提出了一个新的无参数指标——Bcp 指数来识别睡美人文献以缩短重要科学发现的认可时滞[37]。

上述方法中，技术演化法识别突破性创新是对已经发生的突破性创新分析和观察其技术演化过程中的不连续变化。专利分析法也多是利用已经发生的突破性创新对方法进行验证，并且专利引用本身会回避相似专利或者对其创新性造成威胁的专利，因此通过专利的被引识别突破性创新具有一定的局限性。睡美人文献识别突破性创新主要是为了缩短突破性创新成果的认可时滞，但睡美人文献本身即是一种延迟承认现象，该方法无法做到及时预警。利用主题突变识别和网络结构识别突破性创新基于的是科学论文本身的研究内容，分析和研究论文的题名、摘要、关键词、作者等特征项信息。

已有研究大多针对整个中观研究领域开展突破性创新的评估与预测，但随着学科领域划分的日益精细，某个宏观或中观研究领域的突破往往需要多个微观子领域共同促成。由于每个微观子领域的发展进程不同，突破性程度不同，因此有必要针对领域微观主题展开评估与预测。与已有研究相比，本文旨在对微观技术主题的突破性创新进行预测。同时，由于突破性创新在众多技术创新中属于重大创新类型，相对于创新总体数量偏少，因此本研究从新兴技术主题切入，能够增加突破性创新识别的有效性。此外，本研究从推动技术突变的知识创造与扩散的角度出发，追踪微观主题知识的扩散特征，通过观察和提炼不同新兴技术主题所构建的引文曲线的类型和特征识别和预测具有突破性创新潜力的主题，以期能够补充和丰富突破性创新的识别方法。

3 基于引文曲线拟合分析的新兴技术主题的突破性预测方法

3.1 引文曲线的主要类型

引文曲线(citation curve)是指一篇文献发表后的被引次数随时间变化的曲线[37]。A. Avramescu 将引文曲线划分为五种：前三种引文曲线均是先单调递增，后单调递减，根据峰值高度不同，受认可程度也不同；第四种是持续增长型的引文曲线；第五种是在论文发表初期被认可，但突然遭到否定不再被引用的曲线[38]。李江等人通过曲线拟合提出了五种引文曲线：前两种是规则引文曲线，包括经典引文曲线和指数增

长引文曲线；后三种是不规则引文曲线，包括睡美人引文曲线、双峰引文曲线和波型引文曲线[39]，其中规则引文曲线对应于 A. Avramescu 提出的前四种曲线。

本研究在李江等人构建的五种引文曲线分析框架的基础上提出本文可能会出现的四种引文曲线，分别是经典引文曲线、波型引文曲线、双峰引文曲线和指数增长引文曲线（见图 1），并通过曲线拟合的方法对干细胞领域新兴技术主题的引文曲线进行分类和研究。

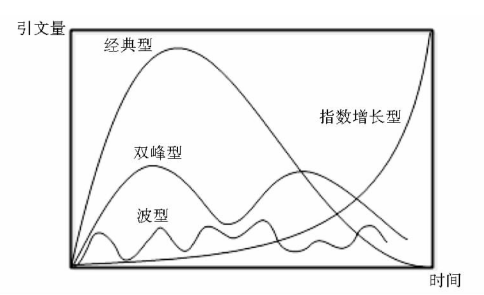


图 1　引文曲线主要类型[39]

3.2　基于不同引文曲线预测新兴技术主题的突破性

本文从新兴技术主题生长变化趋势以及引文曲线对这种趋势的表征能力入手，通过观察引文曲线的规律和特征对新兴技术主题进行突破性预测。研究首先提出应用引文曲线追踪感知新兴技术主题形成突破性创新的三个假设。

3.2.1　引文曲线可表征知识创新的动态过程

创新强度高的新兴技术主题一般会获得科学共同体的较大关注，发文量和引文量都会随着时间的发展明显增加。科学论文是重要的科技创新成果，引文表示学科间知识的流动，引文曲线则在一定程度上表征了科学论文随时间变化的动态价值，也反映了领域知识创造和传播的动态过程。“突变”一词强调变化过程的间断或突然转换[17]，突破性创新的过程同样是动态变化的。具有前瞻性的科学知识和科研成果不仅能够带动科学技术的发展，也是催生重大科学发现和突破性创新的源泉和动力。因此，通过观察新兴技术领域发表的科技论文的被引频次随时间变化的规律来识别突破性创新具有理论可行性。

3.2.2　引文曲线可表征知识跃迁

“跃迁”概念最初来自物理学的量子跃迁理论，指的是原子在吸收光量子后会从低能级的初态跃迁到高能级的终态；反之，从高能级的初态跃迁到低能级的终态时会释放光量子[40]。后来，“跃迁”这一概念被用来隐喻科学技术的重大突破——当科学技术发生重大突破时，其技术轨道往往会呈现飞跃式发展，由此形成新的技术范式和新的技术轨道[41]。突破性创新与渐进性创新不同，其技术轨道是不连续的[14]。技术突破需要创新积累的过程，当大量科学研究活动聚集在某一研究主题上时，该主题可能产生新的科学理论和方法，当该领域的创新累积到一定程度时，技术创新会突破原有轨迹实现跃迁，最终演变为突破性创新。

李柏洲等将知识创造的过程与量子跃迁进行类比，即通过组织学习，知识能级增大，知识系统进入激发态，当新知识产生于知识能级的高能态处时，知识跃迁完成[40]。科学共同体为解决新的科学问题会创造新知识，当新知识足以使原有的知识体系的范式发生根本性变化时，就会实现知识跃迁。同时，科学知识的演进也是不断积累、扩散和跃迁的过程，知识扩散可以用文献间的引用关系来表征，引文曲线最高被引峰值之间的差值能够在一定程度上表示知识的跃迁程度。科学原理的新发现不仅是知识跃迁的重要条件，也是新技术替代旧技术，突破技术瓶颈实现技术升级的关键。因此，引文曲线明显分层，并且实现多次大幅度跃迁的新兴技术主题更可能产生突破性创新成果。

相邻年份最高被引峰值之间的差值越大，引文曲线之间的断层越明显，表明知识的创新程度越高，知识更可能发生质变，呈现跳跃性发展。本研究利用跃迁指数，即相邻年份引文曲线最高被引峰值之间的差值来代表引文曲线的分层程度，这也是判定知识跃迁的标准。

跃迁指数的计算公式为：

$$X_{ik} = \frac{x_{ik} - x_{ik-1}}{x_{ik}^{max} - x_{ik}^{min}} \qquad \text{公式(1)}$$

其中，X_{ik}为跃迁指数，表示第 i 个主题（$i=1,2,3\cdots,n$）第 k 年与第 $k-1$ 年的引文曲线（$k=1,2,3\cdots,m$）最高被引峰值之间的差值，x_{ik}表示第 i 个主题第 k 年的引文曲线的最高被引峰值，x_{ik}^{max}和 x_{ik}^{min}分别是第 i 个主题所有引文曲线最高被引峰值的最大值和最小值。

3.2.3　引文曲线可表征新兴技术主题的发展潜力

新兴技术主题是新颖且呈迅速增长态势的有关科学或者技术问题的研究主题[42]。新颖性不仅是新兴技术主题的主要特征，也是突破性创新的重要特征[4]。新兴技术主题如果能够在近期保持高被引，表明该主题最新创造和生产的科学知识能够迅速获得关注和认可，科学知识的新颖性越高，产生突破性创新的概率也越大。因此，对于新兴技术主题而言，其引文曲线具有持续攀升的特征，则该主题成为突破性创新的潜力更大。

研究通过所有引文曲线的最高被引峰值出现时间

和近 3 – 5 年引文曲线的变化趋势测度科学知识的新颖性，最高被引峰值出现的时间越晚，并且近期保持增长的新兴技术主题，受到的科学共同体的关注度越高，与之相关的科研资源的积累越多，越有可能创造出新的理论与方法，从而催生出突破性创新。

3.3　曲线拟合分析方法

曲线拟合是指通过适当的数学函数或曲线类型来拟合文献的时序被引数据，是研究引文曲线的常用方法。本文对新兴技术主题按照时间切片形成的引文曲线进行对数正态函数拟合，用以判断主题所包含的经典引文曲线类型，对于其他类型的引文曲线则根据折线图识别。

在利用引文曲线追踪感知新兴技术主题形成突破性创新的三个假设的基础之上，本文识别的是新兴技术主题的突破性，该类主题本身是一种新出现且呈现增长态势的主题，但其能否持续发展则受到多种因素的制约。本文从知识跃迁和持续增长两个角度对新兴技术主题的引文曲线进行分析，并提炼出识别突破性创新的主要标准：引文曲线出现明显分层，跃迁次数多，跃迁幅度大；引文曲线在近期保持高被引且持续稳定或快速增长，最高被引峰值出现时间较晚。图 2 为本文利用引文曲线特征识别新兴技术主题突破性的方法模型与分析流程：

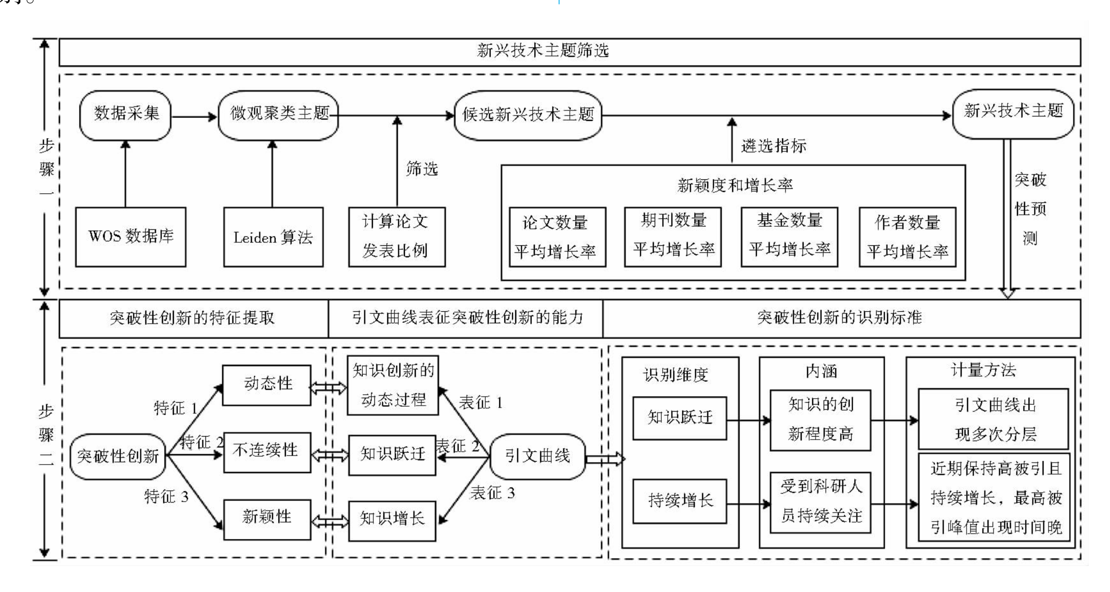


图 2　新兴技术主题突破性识别的方法模型

本文以干细胞领域为例，对该领域的新兴技术主题进行突破性预测。首先，识别新兴技术主题。通过检索领域内的文献总数，根据新兴技术主题的识别指标，即主题的新颖度和增长率，最终筛选得到 26 个新兴技术主题作为突破性预测的分析对象。其次，采集各主题的被引数据构建引文曲线，进行曲线拟合和突破性预测。通过采集 26 个新兴技术主题从 2000 年到 2017 年每一年发文的时序被引数据，构建各主题的引文曲线并进行对数正态函数拟合，结合引文曲线的折线图，对各主题的引文曲线的类型进行分类。最后，提炼不同新兴技术主题引文曲线的主要特征，最终识别和预测有可能发展成为突破性创新的新兴技术主题。

4　实证分析

4.1　数据获取与分析

社区检测是了解大型复杂网络结构的常用方法，本研究通过引文网络发现有意义的研究主题社区。V. A. Traag 等发现 Louvain 算法可能产生任意连接不良的社区，甚至导致社区断开连接，并开发了 Leiden 算法[43]，该算法生成的社区保证是连接的，并且运行速度更快，能发现更好的分区。在 Leiden 聚类算法下，WOS 数据库的所有论文被分为三个层级，宏观层有 22 个聚类，中观层有 868 个聚类，微观层有 4 047 个聚类。本文基于 Leiden 算法采用了微观层聚类主题，并在此基础上进一步分析新兴技术主题。

干细胞是一类具有高度自我更新和多向分化潜能的细胞，为再生医学和疾病治疗提供了新的思路，具有重要的研究价值和发展前景，是生物医学领域的重点研究对象。本文以干细胞领域为例进行实证研究，于 2018 年 10 月 20 日检索到干细胞领域的 422 101 篇论文。将 2004 – 2018 年作为研究年限，同时设定 11 个阶段，422 101 篇文献的微观主题分类的数量随时间切

片的推移呈现增长趋势。为了获取属于干细胞领域的主题,进一步计算了各微观主题包含的干细胞领域论文在全 WOS 数据库中该主题论文集中的比例,删除发表比例低的主题,由此筛选出候选新兴技术主题。

结合新兴技术主题的识别指标,即主题的新颖度和增长率对候选新兴技术主题进一步筛选。新颖度和增长率包含四个指标,分别是各个主题的论文数量、期刊数量、基金数量和作者数量的平均增长率。论文数量表征主题的研究热度,期刊数量表征该主题吸引的投稿数,基金数量表征主题的受资助情况以及国家的重视程度和投入力度,作者数量表征主题受学者的关注程度。

本文选取 11 个阶段中每个阶段论文数量排名前 50 的主题进行计算与分析。新兴技术主题的筛选步骤如下:

第一步,探究各个主题的增长态势。首先,计算每个主题的论文数量在各阶段的增长率,并计算该主题论文数量在全阶段的平均增长率(i=10),把握主题出现的时间和整体发展趋势。此过程中,关注近期(2010年后)出现的主题,而对于近期消失的主题,本文认为其新颖性不足,予以排除,初步筛选出 54 个新兴技术主题。

平均增长率公式为:

$$\overline{V_i} = \frac{1}{i}\sum_{t=1}^{t+i}\frac{N_{t+i}-N_t}{N_t} \qquad 公式(2)$$

其中,i 为时间段,N 为论文数量、期刊数量、基金数量、作者数量,$\overline{V_i}$ 为平均增长率。

第二步,计算与各个主题相关的期刊数量、基金数量、作者数量在全阶段的平均增长率(i=10)。剔除平均增长率为空值的主题,空值表示该主题在期刊、基金或作者数量排名统计中没有进入前 50 名,表明该主题的研究热度或受关注度不够,故予以剔除,比如主题 1297,尽管其论文数量的平均增长率排名靠前,但期刊、基金和作者数量的平均增长率均为空值,说明该主题的研究涉及的作者、期刊以及基金数量的排名均没有进入前 50 名,研究热度不高,故删除该主题。最后剔除 10 个主题,筛选得到 44 个新兴技术主题。

第三步,分析各个主题近期的增长趋势。选择后 5 个阶段作为分析时间窗(i=4),计算剩余主题论文数量在后 5 个阶段的平均增长率。剔除后 5 个阶段平均增长率为负值的主题,负值表明该主题论文数量在近期呈递减趋势,共剔除 18 个主题。最终筛选得到 26 个新兴技术主题,部分主题词如表 1 所示:

表 1　干细胞领域新兴技术主题(部分)

主题编号	干细胞主题标签
1460	stem cell; scaffold; tissue engineering; cell; extracellular matrix; tissue; mesenchymal stem; VITRO; regeneration; decellularized
814	spinal cord; cord injury; stem cell; functional recovery transplantation; spinal cord injury (SCI); mesenchymal stem; neural stem; injury SCI; bone marrow
353	stem cell; cell; surface; mesenchymal stem; extracellular matrix;substrate; differentiation; tissue; hydrogel; biomaterial
161	acute myeloid; myeloid leukemia; stem cell; myelodysplastic syndrome; cell transplantation; patient; AML; hematopoietic stem; leukemia AML; treatment
142	DNA methylation; stem cell; epigenetic; gene; gene expression; EMBRYONIC STEM; expression; human; cell;
2	stem cell; microRNA; miRNA; expression; breast cancer; regulation; function; MESENCHYMAL TRANSITION; cancer cell; MicroRNAs miRNA
469	stem cell; liver fibrosis; expression; progenitor cell; stellate cell; mesenchymal stem; hepatic stellate; hepatocyte; liver disease; bone marrow
261	patient; stem cell; Hodgkin lymphoma; cell transplantation; cell lymphoma; NON HODGKIN; autologous stem; treatment; high dose; relapsed refractory
254	stem cell; mesenchymal stem; stromal cell; bone marrow; MSC; mesenchymal stromal; cells MSC; adipose tissue; expression; derived mesenchymal
648	stem cell; limbal stem; ocular surface; epithelial cell; amniotic membrane; cornea; corneal; expression; transplantation; limbal epithelial

通过上述方法筛选出干细胞领域 26 个新兴技术主题后,本文从 WOS 数据库中采集了每个主题 2000－2017 年每一年的发文数据以及每年发文的时序被引数据,被引数据的统计是自发表当年到 2018 年的被引数。以主题 814 为例,时序被引的部分数据见表 2。

4.2　引文曲线的类型与特征

4.2.1　引文曲线类型

本文利用 Python 进行对数正态函数拟合,并进行 Kolmogorov-Smirnov(K-S)拟合优度检验以判断 26 个新兴技术主题的引文曲线是否属于经典引文曲线,显著性水平取 0.05。剩余的引文曲线类型通过观察折线图的形状和走势判断。

26 个新兴技术主题中共拟合出 467 条引文曲线,其中经典引文曲线有 454 条,占比 97.2%,波型引文曲线有 9 条,占比 1.9%,双峰引文曲线有 1 条,其他引文曲线有 3 条。有 16 个新兴技术主题的引文曲线全部是经典引文曲线,可见经典引文曲线占据主导,其数量远远超过其余类型的曲线(见表 3)。P 值在 0.9 以上的拟合效果好的引文曲线共有 46 条,其中,主题 461 包含的拟合效果好的经典引文曲线数量最多,共计 6 条。

表 2 主题 814 每年发文的时序被引(部分)

发表年	被引年	被引数	发表年	被引年	被引数	发表年	被引年	被引数
2000	2000	7	2000	2013	56	2001	2008	93
2000	2001	58	2000	2014	52	2001	2009	70
2000	2002	110	2000	2015	47	2001	2010	79
2000	2003	87	2000	2016	41	2001	2011	73
2000	2004	112	2000	2017	44	2001	2012	62
2000	2005	88	2000	2018	26	2001	2013	48
2000	2006	89	2001	2001	10	2001	2014	52
2000	2007	73	2001	2002	81	2001	2015	41
2000	2008	69	2001	2003	81	2001	2016	33
2000	2009	51	2001	2004	99	2001	2017	38
2000	2010	53	2001	2005	114	2001	2018	25
2000	2011	49	2001	2006	107			
2000	2012	48	2001	2007	88			

表 3 新兴技术主题的引文曲线类型

主题编号	引文曲线类型					主题编号	引文曲线类型				
	经典	波型	双峰	指数增长	其他		经典	波型	双峰	指数增长	其他
2	17	0	0	0	0	710	18	0	0	0	0
60	17	1	0	0	0	727	18	0	0	0	0
107	18	0	0	0	0	814	18	0	0	0	0
142	16	2	0	0	0	867	18	0	0	0	0
161	18	0	0	0	0	921	15	2	0	0	1
221	18	0	0	0	0	965	18	0	0	0	0
254	17	1	0	0	0	1046	17	1	0	0	0
261	18	0	0	0	0	1095	18	0	0	0	0
353	18	0	0	0	0	1142	17	0	1	0	0
461	18	0	0	0	0	1199	18	0	0	0	0
469	18	0	0	0	0	1290	18	0	0	0	0
581	17	1	0	0	0	1460	18	0	0	0	0
648	16	1	0	0	1	2276	17	0	0	0	1

除经典引文曲线外,波型引文曲线的数量排第二,主题 60、142、254、581、648、921、1046 均包含波型引文曲线。双峰型引文曲线只在主题 1142 中出现,所有主题均未出现指数增长型引文曲线。

单纯引文曲线类型的划分尚不足以辨识新兴技术主题是否具有突破潜力,本文进一步根据突破性预测的方法模型,从知识跃迁和持续增长两个角度出发,观察引文曲线的特征进行突破性预测。

4.2.2 引文曲线特征与突破性创新预测

本文利用跃迁指数、跃迁次数、最高被引峰值出现时间、近期能否保持高被引、近期变化趋势等标准预测新兴技术主题的突破性。根据突破性识别标准,在引文曲线分层的基础上,进一步提炼引文曲线的特征进行突破性预测。

引文曲线知识跃迁的衡量标准包括跃迁指数和跃迁次数。通过计算 26 个新兴技术主题引文曲线的跃迁指数,发现所有引文曲线的跃迁指数的范围集中在 -0.71 到 0.85 之间。结合引文曲线的整体变化幅度,本文将跃迁指数大于 0.14 作为判定引文曲线分层的标准。所有主题的引文曲线均出现分层,跃迁次数为 2 到 6 次不等,跃迁指数最高的是主题 261 于 2012 年的引文曲线发生的跃迁,达到 0.85。新兴技术主题的跃迁指数分布见图 3。

引文曲线持续增长的衡量标准包括:最高被引峰值出现时间晚,近期能够保持高被引,曲线近期变化趋势为增长趋势。近期保持增长是指 2014 年 -2017 年的引文曲线能够在论文发表后的三年内保持增长趋势,由于引文时滞造成无法获取到 2018 年的完整被引记录,因此,引文曲线在 2018 年被引量降低并不代表曲线近期趋势下降。

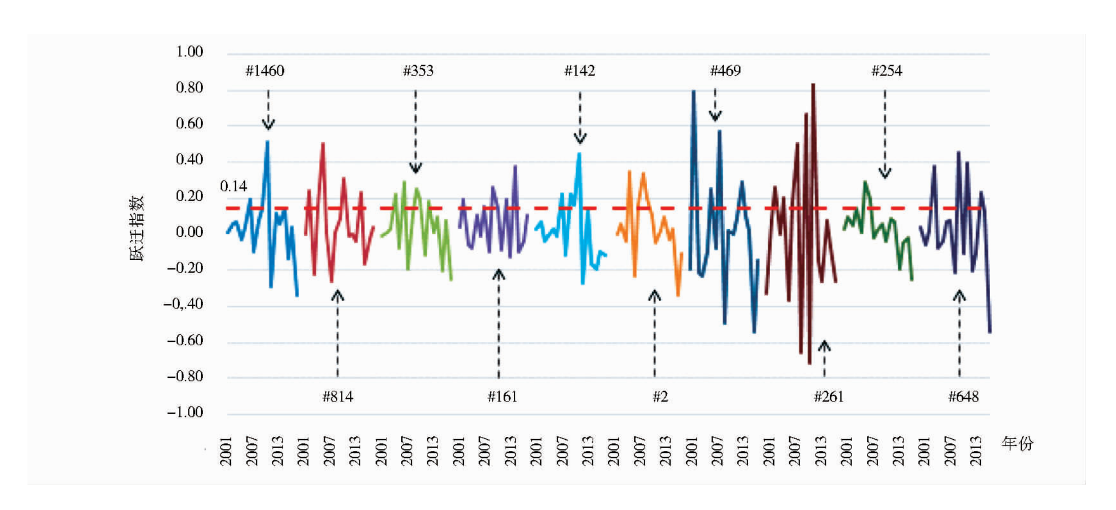


图 3　新兴技术主题的跃迁指数分布

26 个新兴技术主题中，有 11 个主题在未来具有发展成为突破性创新的潜力，有 15 个主题发展成突破性创新的可能性较小。限于篇幅，本文挑选部分代表性主题详细分析。

（1）近期持续增长。不少新兴技术主题的引文曲线近期呈现持续增长的趋势。以主题 1460 为例（见图 4），引文曲线发生了四次跃迁：2000－2005 年，引文量较低，不超过 150 次，曲线变化趋势平稳；2006－2008 年，引文量出现增长，实现第一次跃迁，到达被引峰值后开始缓慢下降；2009 年，引文量突破 300，引文曲线实现第二次跃迁；从 2010 年开始，引文量急剧增长，突破 700 次，引文曲线实现第三次跃迁，跃迁指数高达 0.51；2014 年引文曲线再次跃迁。所有引文曲线的最高被引峰值出现时间较晚，为该主题 2014 年发表的论文在 2018 年的被引，引文曲线在近期能够保持高被引和持续增长的态势。这些特征表明该主题产出的科学知识不仅跃迁次数多，跃迁幅度大，并且正在受到科学共同体的关注，具有成为突破性创新的潜力。主题 1460 研究的是脱细胞干细胞在组织再生医学中的应用。2010 年 9 月，英国利兹大学的 J. Fisher 和 E. Ingham 教授通过死亡捐赠者或动物的部分身体制作裸露支架，成功利用患者的干细胞制成新的身体部件。同年 10 月，美国维克森林大学医学院主管兼教授 S. Soker 团队使用人体干细胞制造出微型人体肝脏。该项技术保留了组织或细胞特异性的 ECM 蛋白组分和微结构，剥离了触发排异反应的材料，得到广泛应用[44]。2016 年，J. P. Guytte 等用人的诱导多能干细胞作为种子细胞接种到器官捐赠者的心脏脱细胞支架构建了组织工程心脏[45]。

主题 814（见图 5）的引文曲线经历了“引文量低，

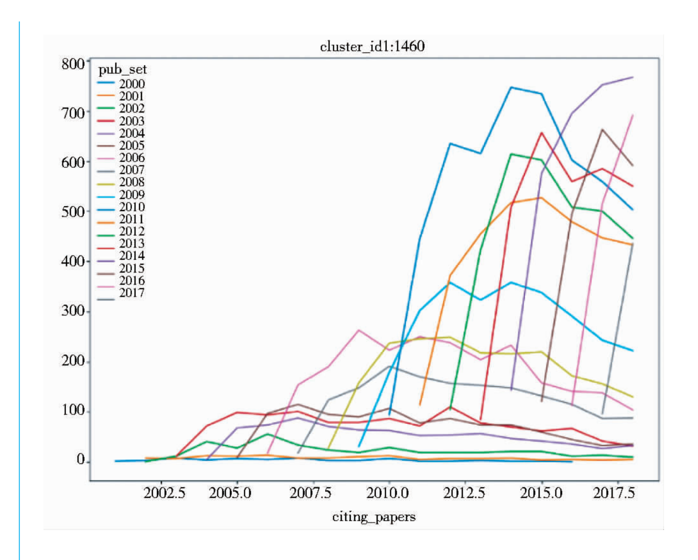


图 4　主题 1460 引文曲线

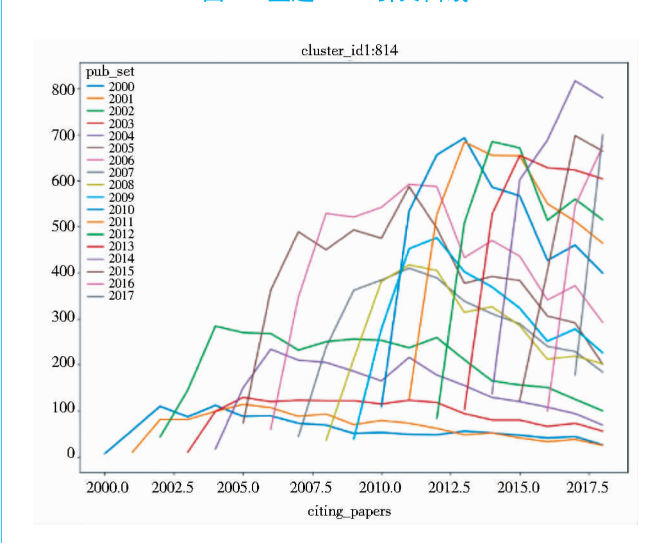


图 5　主题 814 引文曲线

趋势平稳——引文量较高，波动式缓慢下降——引文量高，波动式快速下降——引文量很高，持续攀升”四个阶段的发展，引文曲线发生五次跃迁，近期发表的论

文能够保持高被引且近期的引文曲线仍在快速增长，最高被引峰值为2014年的论文在2017年的被引，出现时间较晚，表明该主题近年发表的论文迅速获得了同行的认可，论文中的理论与方法为干细胞领域的发展做出了重要贡献，有较大潜力产生突破性创新成果。主题814研究的是利用干细胞治疗脊髓损伤。根据ClinicalTrials. gov的记载，目前国际上关于干细胞治疗脊髓损伤的临床试验已有50例，涉及骨髓干细胞、间充质干细胞、自体干细胞移植等多个研究方向[46]。2018年5月，E. Curtis等发布了利用神经干细胞治疗脊髓损伤的新的研究成果，4名受试者中有3名患者的病症得到显著改善[47]。2018年8月，K. Hiromi等在*Nature*子刊上发表研究成果，通过人类多能干细胞(hPSCs)创造了脊髓源性神经干细胞，有望用于脊髓损伤的临床治疗[48]。

与主题814类似，主题353的引文曲线也大致分四个阶段(见图6)。二者区别在于，主题353在第二阶段的引文曲线呈波动式上升趋势，第三阶段引文曲线的下降速度也较为缓慢，第四阶段被引频次持续增长，可看作是经典引文曲线的单调递增部分。整体上，主题353的引文曲线分层明显，共跃迁五次，最高被引峰值出现时间较晚，也出现在2014年发表的论文在2017年的被引，表明主题353引文曲线的生命周期较长，近期发表的论文能够保持较高被引且引文量持续攀升，研究成果的影响更持久，持续获得同行的关注和认可，有可能产生突破性创新。主题353研究的是细胞外基质可以调节干细胞的命运。2018年9月，W. J. Sullivan等在*Cell*杂志上发表研究成果，发现细胞外基质可以通过调节细胞的糖代谢控制细胞在体内的运动，该研究对于揭示癌细胞转移有重要意义[49]。

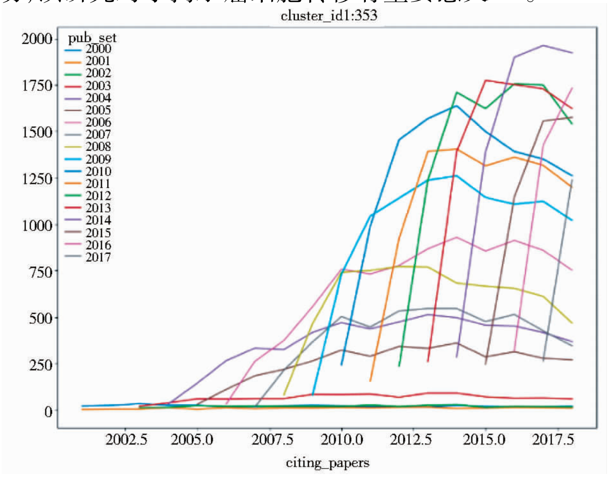


图6 主题353引文曲线

主题161(见图7)的引文曲线大致分为三个阶段。第一阶段的引文曲线在论文发表2-3后便开始快速上升，之后呈现波动式的缓慢下降趋势；第二阶段的引文曲线在2009年实现跃迁，之后依旧呈现波动式下降趋势；第三阶段的引文曲线在2014年完成跃迁，最高被引峰值出现在2014年。该主题的引文曲线跃迁次数较多，有6次，并且近期能保持相对较高的被引量，引文曲线呈上升趋势，表明该主题研究热度较高，受到研究人员的持续关注，有望发展成突破性创新。主题161研究的是造血干细胞的临床应用，目前越来越多的患者正在接受异基因造血干细胞移植来治疗急性髓性白血病。

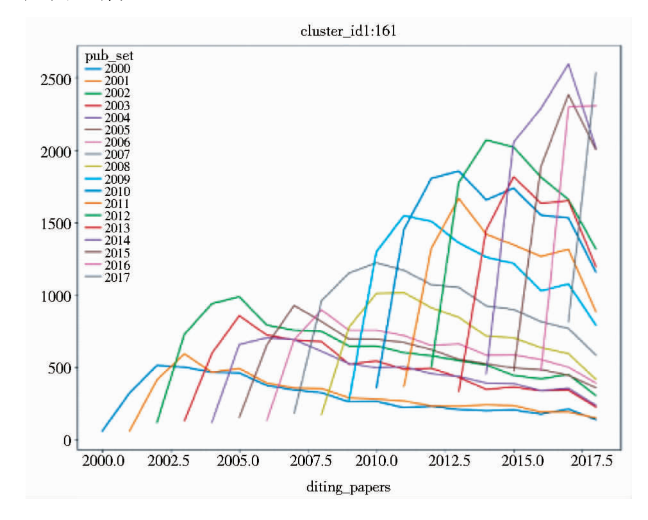


图7 主题161引文曲线

(2)近期持续下降。某些新兴技术主题的引文曲线尽管发生多次跃迁，但近期的研究热度下降。以主题142为例(见图8)，引文曲线共跃迁五次，大致分三个阶段：2000-2006年的引文量大致在500以内，引文曲线近似于一条直线，表明该主题前期发表的论文在之后的十几年中引文量没有显著变化；从2007年到2010年，该主题的引文曲线共完成三次跃迁，引文量进入快速增长阶段，主题的受关注程度开始攀升，该阶段的引文曲线均在论文发表3-4年后呈现缓慢的下降；第三阶段的引文曲线属于经典引文曲线，从2011年开始引文数量急剧增长，完成第四次大幅度的跃迁，该阶段所有引文曲线均在论文发表第二年后达到峰值，之后又快速下降。该主题所有引文曲线的最高被引峰值出现时间较早，在2011年，近期引文曲线下降趋势明显且未保持较高的被引，表明该主题尽管前期得到了广泛关注，科学知识发生了跃迁，但近几年的研究热度相较之前有所下降，可能遭遇技术瓶颈，近期内成为突破性创新的可能性较小。主题142研究的是干细胞中的DNA甲基化。DNA

甲基化是表观遗传机制中一种重要的调控方式，可用于胚胎干细胞、诱导多能干细胞和间充质干细胞等的谱系分类和质量控制。领域专家指出，由于 DNA 甲基化属于机制研究，不太可能演化为突破性创新。

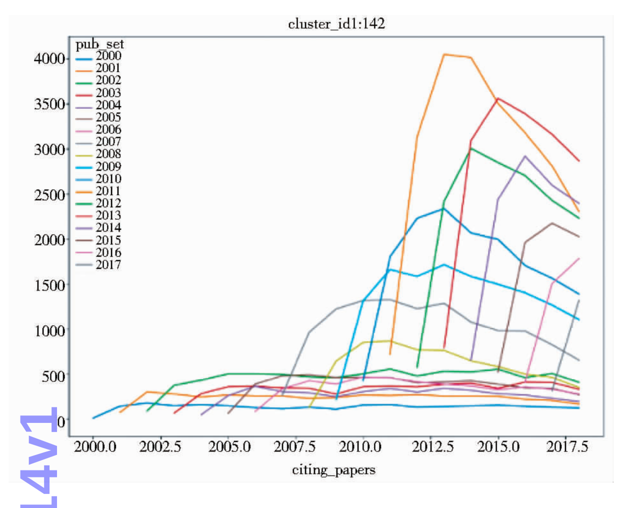


图 8　主题 142 引文曲线

主题 2 的引文曲线发生四次跃迁，其发展大致分为三个阶段(见图 9)：第一阶段是 2001－2004 年，引文量较低，均在 500 以内，引文曲线很平稳，近似于一条直线；第二阶段是 2005－2007 年，引文量有所增长，之后开始波动式缓慢下降；第三阶段是 2008－2017 年，引文量激增，到达峰值后开始波动式快速下降。该主题的引文曲线跃迁次数较多，跃迁幅度较大，但近期引文曲线整体呈下降趋势，近两年引文量虽有所增长，但与巅峰期相比差距明显，表明该主题的研究热度不及之前，不太可能演化成突破性创新。主题 2 研究的是小分子 RNA 在干细胞的调控中发挥着重要作用。小分子 RNA 是一类调控因子，仍属于作用机理研究，产生突破性创新的可能性较小。

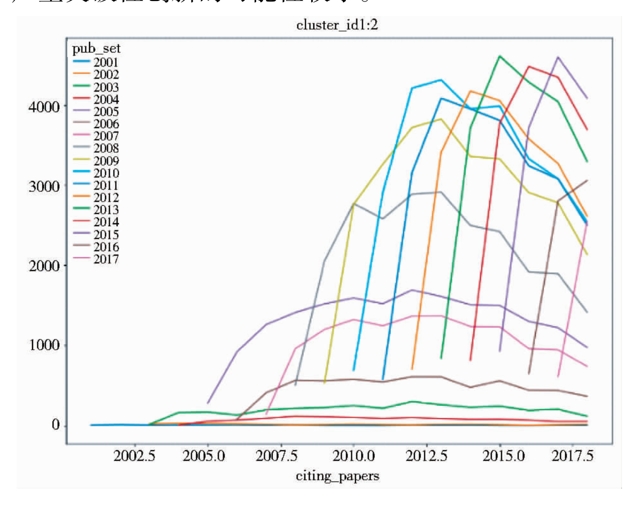


图 9　主题 2 引文曲线

(3)生命周期短。引文曲线生命周期较短的主题表现为引文量快速增长达到峰值后又迅速下降。以主题 469 和主题 261 为例(见图 10、图 11)，这两个主题的论文一经发表，就迅速获得认可并得到较高的引用，且普遍在发表 2－3 年后就达到被引峰值，但之后这些论文并未保持较高的被引，引文量又迅速回落至较低水平。表明这两个主题的论文受关注时间早，被引量上升快，下降也快，引文曲线的生命周期普遍较短，虽然对本领域的科学研究做出过贡献，但没有保持持久的影响力。主题 469 和主题 261 的引文曲线分别发生 4 次跃迁和 6 次跃迁，最高被引峰值分别出现在 2015 年和 2012 年，两个主题的引文曲线在近期均未能保持高被引和增长趋势。创新程度高的新兴技术主题一般会保持较高的关注度和研究热度，因此这两个主题近期发展为突破性创新的可能性较小。主题 469 研究的是干细胞在肝再生中的应用。目前用于肝病治疗的干细胞有肝脏干细胞、人类胚胎干细胞、诱导多能干细胞等[50]，但干细胞对肝脏再生机制的基础研究仍在探索阶段，不太可能产生突破性创新。主题 261 研究的是利用干细胞治疗霍奇金淋巴瘤。大剂量化疗序贯自体造血干细胞移植已成为复发/难治霍奇金淋巴瘤的标准治疗方案[51]，该技术已相对成熟，近期再产生突破性创新的可能性较小。

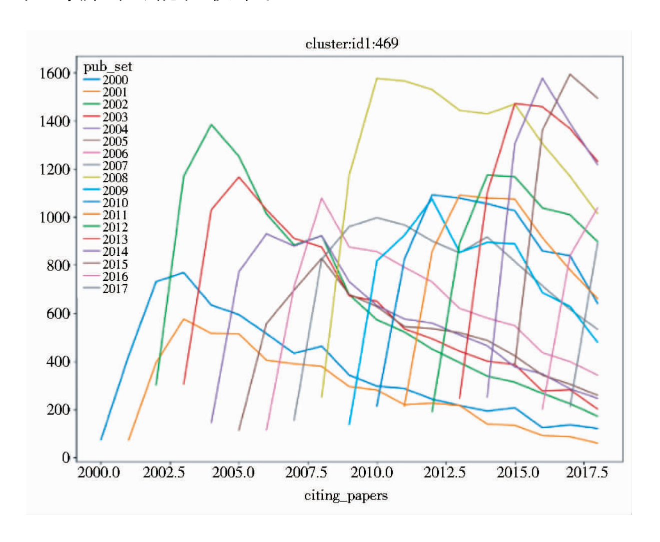


图 10　主题 469 引文曲线

(4)相近年份引文曲线变化趋势一致。某些主题的引文曲线在相近年份表现出了相近的增长和下降速率。以主题 254 为例(图 12)，2004 年与 2005 年的引文曲线几乎重合，2006－2011 年间的 6 条引文曲线的形状也非常相似，2011－2013 年，这 6 条引文曲线都开始攀升，2013－2014 年被引频次逐渐回落，2014－2015

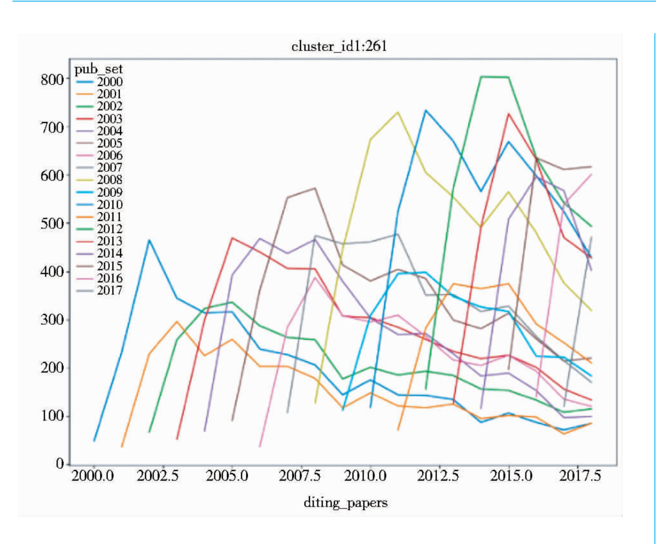


图 11 主题 261 引文曲线

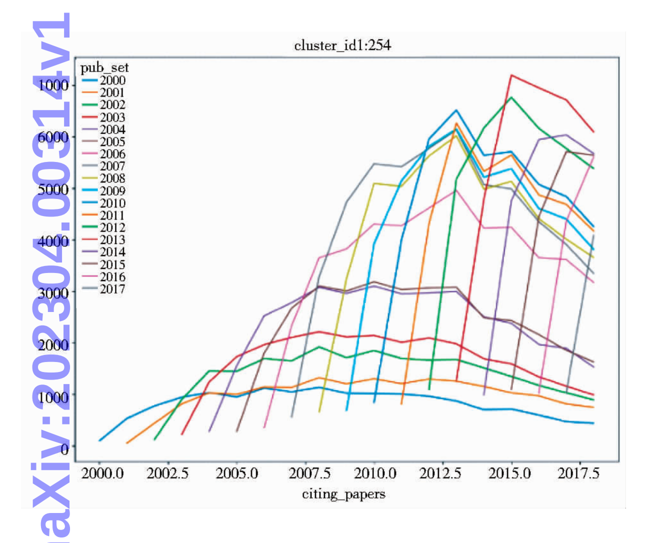


图 12 主题 254 引文曲线

年，曲线又开始回升，之后逐渐下降。相近年份的引文曲线除了突变方式一致，即均在相同年份发生单调性的变化，变化速率也较为接近。文献发表的时刻以及发表时所处社会环境特征是引文扩散中的时间环境要素[52]，因此这种现象可能缘于该主题在不同年代发表的论文所处的科研环境不同，但在相近年份发表的论文则处于相似的环境中，受到的科研关注度也较为接近，因此引文曲线表现出近乎一致的变化趋势。

主题 648 的引文曲线（图 13）也具有相近年份变化趋势一致的特征，同一阶段的引文曲线形状较为相似。这种变化趋势的相似性是否对于预测近几年引文曲线的发展走向有参考价值还有待后续观察。值得一提的是，主题 254 与主题 648 在近几年发表论文的被引频次仍在以较快的速度持续增长，甚至有可能突破之前的被引高峰，创造新的峰值，表明这两个新兴技术主题在干细胞领域的研究热度不减，研究成果仍被领

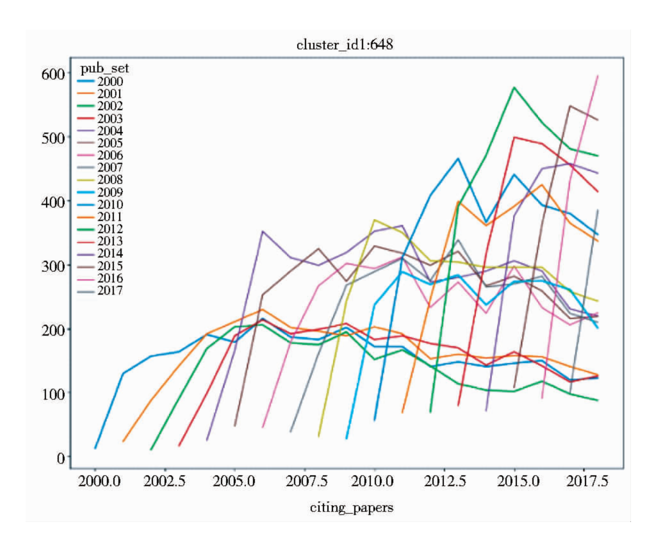


图 13 主题 648 引文曲线

域内的同行专家广泛关注，具有发展成为突破性创新的潜力。主题 254 是间充质干细胞及其应用研究。截止到 2019 年 3 月，我国共有 38 个干细胞临床研究项目通过备案，其中有 29 个临床项目使用了间充质干细胞[53]。主题 648 是角膜缘干细胞及其应用研究。利用患者自体细胞或羊膜细胞，角膜缘干细胞移植技术能够成功修复功能性角膜上皮。自体角膜缘干细胞扩增培养细胞治疗已经在欧盟获得批准使用[54]。该技术具有广阔的应用场景。

5 讨论与结论

本文提出利用引文曲线追踪感知新兴技术主题形成突破性创新的三个假设，即引文曲线可表征知识创新的动态过程、可表征知识跃迁、可表征新兴技术主题的发展潜力，从知识跃迁和持续增长两个维度出发，总结识别突破性创新的标准，提出跃迁指数的概念和计量指标，并构建了基于引文曲线识别新兴技术主题突破性的方法模型。研究利用每个新兴技术主题发文的时序被引数据构建引文曲线并进行曲线拟合，对不同新兴技术主题包含的引文曲线的类型进行分类，并根据突破性创新的识别标准提炼引文曲线的主要特征进行突破性预测。研究对引文曲线拟合特征和预测结果与领域专家的解读基本吻合。

新兴技术主题的引文曲线主要包括四个特征：近期持续增长；近期持续下降；生命周期短；相近年份变化趋势一致。四类特征中，近期持续增长的引文曲线如果有明显分层、跃迁次数多、跃迁幅度大，最高被引峰值出现时间较晚，并且能够在近几年保持高被引的新兴技术主题由于其研究成果的影响力更长远，受关

注度和研究热度较高,更有可能在未来产生重大的突破性创新成果;近期持续下降的引文曲线其所在的新兴技术主题尽管实现了科学知识的跃迁,但研究热度正在下降,不太可能演化为突破性创新;生命周期短的引文曲线引文量迅速上升又快速回落,这类主题虽然受关注的时间早,但研究热度下降较快,研究成果不能保持长久的影响力,成为创新强度高的突破性技术的可能性较小;另外,有些主题的引文曲线在相近年份形状极为相似,可能是由于年份越接近,这些主题所处的科研环境较类似,因此引文轨迹的突变方式和变化速率也更为接近。

由于突破性创新的复杂性,难以有统一的指标实现对各个微观主题的突破性预测。研究发现,跃迁指数、跃迁次数和最高被引峰值出现时间三个识别标准仅可作为参考,跃迁次数多或最高被引峰值出现时间晚的主题并不一定会发生突破性创新。引文曲线近期保持高被引和近期持续增长是预测突破性创新的主要标准,同时具备这两个条件的主题更具发展成突破性创新的潜力,因此,单一标准不足以描述整个主题的发展,多个标准综合评估可提高预测分析的准确性。

突破性创新的识别及预测一直是科技情报分析的核心工作。利用科技文献的引文曲线识别新兴技术主题的突破性,为突破性创新识别提供了新的情报分析思路,丰富和扩展了突破性创新预测的情报分析方法,可为领域专家和规划管理人员提供有效的情报支撑。结合领域专家对实证结果的专业评估以及干细胞领域相关主题的研究进展和科学突破,证明利用引文曲线可以自动、高效追踪感知新兴技术主题的突破性,识别结果可以辅助领域专家和决策咨询者更好地制定领域发展规划。

本研究还存在一些不足,从引文曲线的角度对新兴技术主题进行突破性预测是基于论文的被引数据,但论文被引具有时滞性,还会受到数据库入库时间等其他因素的影响,存在一定的误差。未来考虑通过设置引文时间窗和统计篇均被引的方式减少误差。同时,鉴于突破性创新预测的复杂性,未来还需要进一步结合不同新兴技术主题相关的科技政策布局等进行辅助分析和预测,以尽早识别出具有突破性潜力的新兴技术主题。

参考文献:

[1] 张洪石. 突破性创新动因与组织模式研究[D]. 杭州: 浙江大学, 2005.
[2] ROTOLO D, HICKS D, MARTIN B R. What is an emerging technology? [J]. Research policy, 2015, 44(10):1827－1843.
[3] 罗瑞, 许海云, 董坤. 领域前沿识别方法综述[J]. 图书情报工作,2018,62(23):119－131.
[4] 张金柱, 张晓林. 基于专利科学引文的突破性创新识别研究述评[J]. 情报学报,2016,35(09):955－962.
[5] 万宁. 浅析颠覆性创新、破坏性创新和突破性创新三者关系[J]. 商, 2015(30):122－123.
[6] 陈傲, 柳卸林. 突破性技术从何而来? ——一个文献评述[J]. 科学学研究,2011,29(9):1281－1290.
[7] DESS G G, BEARD D W. Dimensions of organizational task environments[J]. Administrative science quarterly, 1984, 29(1):52－73.
[8] KOSHLAND D E. The cha-cha-cha theory of scientific discovery[J]. Science, 2007, 317(5839): 761－762.
[9] TUSHMAN M L, ANDERSON P. Technological discontinuities and organizational environments[J]. Administrative science quarterly, 1986, 31(3):439－465.
[10] KOTELNIKOV V. Radical innovation versus incremental innovation[M]. Boston: Harvard Business School Press, 2000.
[11] DOYLE J F. Radical innovation: how mature companies can outsmart upstarts[J]. Research-technology management, 2000, 43(10):706－707.
[12] ZHOU K Z, YIM C K, TSE D K. The effects of strategic orientations on technology-and market-based breakthrough innovations[J]. Journal of marketing, 2005, 69(2):42－60.
[13] CHANDY R K, TELLIS G J. Organizing for radical product innovation: the overlooked role of willingness to cannibalize[J]. Journal of marketing research, 1998, 35(4):474－487.
[14] 付玉秀, 张洪石. 突破性创新:概念界定与比较[J]. 数量经济技术经济研究, 2004, 21(3): 73－83.
[15] SCHOENMAKERS W, DUYSTERS G. The technological origins of radical inventions[J]. Research policy, 2010, 39(8):1051－1059.
[16] DAHLIN K B, BEHRENS D M. When is an invention really radical?: defining and measuring technological radicalness[J]. Research policy, 2005, 34(5):717－737.
[17] 张金柱. 利用被引科学知识的突变识别突破性创新[M]. 北京: 科学出版社, 2017.
[18] SMALL H. Referencing as cooperation or competition[M]. Berlin: Walter de Gruyter, 2016.
[19] YOON J, KIM K. Identifying rapidly evolving technological trends for R&D planning using SAO-based semantic patent networks[J]. Scientometrics, 2011, 88(1):213－228.
[20] SOOD A, TELLIS G J. Technological evolution and radical innovation[J]. Journal of marketing, 2005, 69(3):152－168.
[21] ANDERSON P, TUSHMAN M L. Technological discontinuities and dominant designs: a cyclical model of technological change[J]. Administrative science quarterly, 1990, 35(4):604－633.
[22] TELLIS G J. Disruptive technology or visionary leadership? [J].

Journal of product innovation management, 2006, 23(1):34 -38.
[23] KUHN T S. The structure of scientific revolution[M]. Chicago: University of Chicago Press, 1999.
[24] 李勇, 安新颖, 赵迎光, 等. 结合知识组织体系的突发主题监测研究[J]. 情报理论与实践, 2013, 36(5):120 -123.
[25] KLEINBERG J. Bursty and hierarchical structure in streams[J]. Data mining & knowledge discovery, 2003, 7(4):373 -397.
[26] 张金柱, 张晓林. 利用引用科学知识突变识别突破性创新[J]. 情报学报, 2014, 33(3): 259 -266.
[27] 王莉亚. 基于离群数据的主题演化研究[D]. 北京: 中国科学院研究生院, 2012.
[28] ARTHUR W B. The structure of invention[J]. Research policy, 2007, 36(2):274 -287.
[29] ARTHUR W B. The nature of technology: what it is and how it evolves[M]. New York: Free Press, 2009.
[30] FUNK R J, OWEN-SMISH J. A dynamic network measure of technological change[J]. Management Science, 2016, 63(3):587 -900.
[31] CHEN C. Turning points: the nature of creativity[J]. Nature of creativity contemporary psychological perspectives, 2011, 18(1): 87 -98.
[32] CHEN C, CHEN Y, HOROWITZ M, et al. Towards an explanatory and computational theory of scientific discovery[J]. Journal of informetrics, 2009, 3(3):191 -209.
[33] CAMPANARIO J M. Rejecting and resisting Nobel class discoveries: accounts by Nobel laureates[J]. Scientometrics, 2009, 81(2):549 -565.
[34] FANG H. An explanation of resisted discoveries based on construal-level theory[J]. Science & engineering ethics, 2015, 21(1): 41 -50.
[35] RAAN A F J V. Sleeping beauties in science[J]. Scientometrics, 2004, 59(3):467 -472.
[36] PALOMERAS N. Sleeping patents: any reason to wake up? [J]. Iese research papers, 2003, 20(35):D506.
[37] 杜建. "睡美人"文献的识别方法与唤醒机制研究[D]. 南京: 南京大学,2017.
[38] AVRAMESCU A. Actuality and obsolescence of scientific literature [J]. Journal of the American Society for Information Science & Technology, 2010, 30(5):296 -303.
[39] 李江, 姜明利, 李玥婷. 引文曲线的分析框架研究——以诺贝尔奖得主的引文曲线为例[J]. 中国图书馆学报, 2014, 40(2):41 -49.
[40] 李柏洲, 赵健宇, 苏屹. 基于能级跃迁的组织学习 - 知识创造过程动态模型研究[J]. 科学学研究,2013,31(6):913 -9225.
[41] 张立超, 刘怡君. 技术轨道的跃迁与技术创新的演化发展[J]. 科学学研究,2015,33(1):137 -145.
[42] 郭涵宁. 多元科学指标视角下的新兴研究领域识别探索[D]. 大连: 大连理工大学,2013.
[43] TRAAG V A, WALTMAN L, VAN ECK N J. From Louvain to Leiden: guaranteeing well-connected communities[J]. Scientific reports, 2019, 9(1):5233.
[44] 生物谷. 【盘点】再生医学中干细胞和新材料的研究和应用(一)[EB/OL]. [2019 - 09 - 16]. https://meeting. bioon. com/2018StemCells/news-detail/b04d4462ec057124.
[45] GUYETTE J P, CHAREST J M, MILLS R W, et al. Bioengineering human myocardium on native extracellular matrix[J]. Circulation research, 2016, 118(1):56 -72.
[46] ClinicalTrials. gov[EB/OL]. [2019 -09 -16]. https://clinicaltrials. gov/ct2/results? cond = stem + cell + cord + injury&term = &cntry = &state = &city = &dist = .
[47] CURTIS E, MARTIN J R, GABEL B, et al. A first-in-human, phase I study of neural stem cell transplantation for chronic spinal cord Injury[J]. Cell stem cell, 2018, 22(6):941 -950.
[48] HIROMI K, KEN K, ADLER A F, et al. Generation and post-injury integration of human spinal cord neural stem cells[J]. Nature methods, 2018, 15(9):723 -731.
[49] SULLIVAN W J, MULLEN P J, SCHMID EW, et al. Extracellular matrix remodeling regulates glucose metabolism through TXNIP destabilization[J]. Cell, 2018, 175(1):117 -132.
[50] 国际肝病. APASL2019 焦点丨王福生院士:干细胞治疗肝硬化和肝衰竭——进展与研究热点[EB/OL]. [2019 - 09 - 16]. http://www. ihepa. com: 8088/default/htmlDocument/2019-02-22/detail_16360. html.
[51] 王华. 复发/难治性霍奇金淋巴瘤自体造血干细胞移植前苯达莫斯汀、吉西他滨、长春瑞滨作为挽救化疗方案较好[EB/OL]. [2019 -09 -16]. https://www. bjcancer. org/Mobile/Article/Index/4091.
[52] 闵超, DING Y, 李江, 等. 单篇论著的引文扩散[J]. 情报学报,2018,37(04):341 -350.
[53] 优竹网. 趋势丨 37 个干细胞临床项目通过国家备案,治疗疾病范围极广(附全目录)[EB/OL]. [2019 -09 -16]. http://www. sohu. com/a/317270049_827862.
[54] 亿欧. 干细胞治疗研究进展和产业分析[EB/OL]. [2019 -09 -16]. https://www. iyiou. com/p/95934. html.

作者贡献说明:

曹艺文:参与讨论研究思路,负责文献调研、撰写和修改论文;

许海云:确定选题,提出研究思路,指导论文修改;

武华维:参与讨论研究思路,负责数据处理与作图;

罗瑞:参与新兴技术主题识别。

曹艺文，许海云，武华维，等.基于引文曲线拟合的新兴技术主题的突破性预测——以干细胞领域为例[J].图书情报工作,2020,64(5):100-113.

Study on Radical Innovation Prediction to Emerging Technology Topics Based on Citation Curve Fitting ——Taking the Field of Stem Cells as an Example

Cao Yiwen[1,2] Xu Haiyun[1,3] Wu Huawei[1,2] Luo Rui[1,2]

[1] Chengdu Documentation and Information Center, Chinese Academy of Sciences, Chengdu 610041

[2] Department of Library, Information and Archives Management, School of Economic and Management, University of Chinese Academy of Sciences, Beijing 100190

[3] Institute of Scientific and Technical Information of China (ISTIC), Beijing 100038

Abstract: [**Purpose/significance**] Through the fitting analysis of the citation curves of emerging technology topics, this paper refined and summarized the main types and characteristics of the citation curves, in order to provide a useful reference for the study of prediction methods for radical innovation topics at the micro level. [**Method/process**] Firstly, the hypothesis of using citation curves to trace the emerging technology topics to produce radical innovation was proposed and the concept and measurement method of the transition index were proposed. The criteria of predicting radical innovation were summarized and the method model of radical innovation identification was constructed from the two dimensions of knowledge transition and continuous growth. Then used the time slices cited data of each emerging technology topic to construct the citation curves, classify and summarize the types and characteristics of citation curves of different emerging technology topics. [**Result/conclusion**] The citation curves of emerging technology topics includes four types: continuous growth in the near future; continuous decline in the near future; short life cycle; consistent trends in similar years. According to the identification criteria of radical innovation, the citation curves have the potential to become radical innovation if they have many transitions and a large transition range, and the highest citation peak occurs later, and maintains a high citation rate in the near future with a sustained stable or rapid growth. Combined with the assessment of the prediction results by domain experts and the research progress of different emerging technology topics, it is verified that the citation curve fitting analysis can effectively predict the breakthrough of emerging technology topics.

Keywords: radical innovation technical forecast citation curve emerging technology topic